

Validating the importance of criteria for assessing climate change scenarios

Bojan Srdjevic, Zorica Srdjevic and Milena Lakicevic

ABSTRACT

While evaluating climate impacts within different climate change scenarios, analysts and stakeholders may have different goals and therefore it is usually difficult to define a common decision-making framework applicable for various practical uses. In this paper, we combine two different group decision-making methodologies to prioritise criteria for assessing output information from regional climate models. The first is based on use of a multi-criteria analytic hierarchy process (AHP) to determine weights of criteria as cardinal information about their importance. The second methodology uses two voting methods, namely Borda count and Approval voting, to generate ordinal information (ranks) for criteria. A set of five criteria is assessed by 16 PhD students from the field of climatology, and generated decisions about their importance in the validation of regional climate models' quality are summarised, compared, and critically discussed. The paper is closed with recommendations for further research.

Key words | AHP, approval voting, borda count, climate change scenario, evaluation criteria, group decision-making

Bojan Srdjevic (corresponding author)

Zorica Srdjevic

Faculty of Agriculture, Department of Water Management, Group for Systems Analysis and Decision Making,
University of Novi Sad,
Trg D. Obradovica 8, Novi Sad 21000,
Serbia
E-mail: bojans@polj.uns.ac.rs

Milena Lakicevic

Faculty of Agriculture, Department of Fruit Science, Viticulture, Horticulture and Landscape Architecture,
University of Novi Sad,
Trg D. Obradovica 8, 21000 Novi Sad,
Serbia

INTRODUCTION

Uncertainties in climate changes

Addressing climate information at the regional level can be done in many ways, according to the work of researchers and practitioners reported in the literature. For instance, there are many reviews of downscaling work, and discussions on the relative merits and limitations of the different techniques (Giorgi & Mearns 1999; Wang *et al.* 2004; Laprise *et al.* 2008; Feser *et al.* 2011; Themeßl *et al.* 2012; Maraun *et al.* 2015). An interesting discussion of dynamical downscaling (DD) and statistical downscaling (SD) techniques is given in Giorgi *et al.* (2009). While the DD techniques are more related to global and regional climate models (GCM, RCM), SD techniques are primarily based on statistical relationships between large-scale predictors and regional-to-local scale predictands, which are then applied to the output from climate model simulations.

A good example of assessing uncertainties in a regional climate scenario is presented in Seneviratne (2012), where two major steps are proposed. In the first step, the confidence level of scenarios should be determined as low, medium, or high. If it is high, one should proceed with the second step, and perform a likelihood assessment to determine levels such as: virtually certain, very likely, likely, more likely than not, and about as likely as not. Definition of confidence levels in assessing changes in climate on a global scale can also be found in Seneviratne (2012).

Uncertainty is difficult to express and measure, and even harder to control and reduce accordingly, due to its inherent nature and complexity. As stated in Hallegatte *et al.* (2012; pp. 5–6), 'the climate change uncertainties arise from three major sources: (1) inability to predict emissions of greenhouse gases; (2) limited scientific knowledge of the functioning of the climate system and of affected systems

and (3) imperfect understanding of natural variability, i.e. the global climate dynamics, linked to the chaotic behaviour of the climate system. These three uncertainties are sometimes referred to as policy, epistemic, and aleatory uncertainty, respectively. Their respective shares depend on the timescale and the space scale; that is, at global scale and over the short term, natural variability and model response play the largest roles, and the emission a very small role. Over the long term, the emissions dominate other sources of uncertainty.'

Complex mathematical models are developed to manage uncertainty properties such as long-term reliability (risk), resilience, vulnerability, etc. In most cases, they are computerised and represented by programs for simulating climate processes, predicting the future, performing automatic downscaling or upscaling, or creating input to models for specific purposes in different sectors; e.g. for simulating the growth of plants and computing food production in a given growing season, with related economic effects such as the balance of benefits and costs. Models are commonly integrated within software systems, mainly with database management systems, decision support systems, and the geographic information system.

Past evidence suggests that our ability to describe and 'manage' uncertainties and predict the future is rather limited, and that these limitations need to be considered in the way decisions are made. Decision-makers have always used various decision-making methodologies to manage future uncertainty. Due to developments in theory and practice supported by advances in computer-based technologies, methodologies nowadays include those based on artificial intelligence, fuzzy sets, and multilayer (hierarchical) decision architectures.

Still, it is very difficult to make decisions incorporating the complexity of climate change processes and related uncertainties. In such circumstances, it is essential to select a proper decision-making approach, as well as supporting tools capable of managing large differences between climate models, downscaling and upscaling methods. Thus, this paper aims to identify one possible group decision-making approach to detect the importance of criteria (attributes) which ensure the building of a useful and valid framework for the assessment of different climate scenarios, with a particular focus on climate uncertainty.

Climate scenarios, systems approach, and decision-making

Climate scenarios can be analysed in different ways and by different means. Where decision-making is included as a part of the evaluation process, decision-makers must be aware of the following: (1) most decisions must balance multiple criteria; (2) a structured approach is required; (3) information can be complex, changing, and conflicting; and, above all, (4) no right answers exist in a mathematical sense.

Whenever decision-making tools are used, important adaptation characteristics and requirements should be respected. Adaptation frameworks must be established because different parts of the adaptation process exist in many aspects, such as dynamics in time, distribution in space, and organisation (e.g. from a manipulation point of view) (Liu *et al.* 2008). For instance, decision-making processes may differ between decision-makers in different places; or they can be positively, negatively, or neutrally led to elicit trustworthy (valid) judgments about the relative importance of decision elements. Furthermore, an adaptation is an internally generated response of a decision-making process as a closed system to external forces such as various influences, noises, lack of information, unwillingness of humans to participate, or their incompetence. Decision-making can be a continuous process, but it can also be intermittent. It can be divided into phases or be repeatable, as in the case presented in this paper where we organised the process into two separate sessions with a break of three hours, to allow the participants time to slightly forget about their previous judgments and possibly modify their opinions, or to participate in a brainstorming part of the decision process as members of other groups and/or sub-groups.

From a systems theory point of view, the decision-making process has several generic steps. Its 'systems analysis' section is designed to provide for the adaptation of the steps to a particular problem, e.g. to define what to use and what to ignore. Conceptually, a tracking back is required if one wishes to trust the decision-making process itself, and especially its outcomes. A systems approach is embedded into the decision-making process in a way that provides feedback, which is an important adaptation effect. Monitoring of the outcomes of various stages in the decision-making process ensures an assessment of whether

the outcomes of some stages are as expected. If the feedback works properly, the adaptation is effective and the whole process is observable and controllable at the same time. The feedback loop can be partly reversible (so that some previous stages are repeated before the process proceeds), and the feedback effect can be added to a repertoire of adaptive options. If feedback does not work, evaluation is needed regarding what went wrong and why, and the number of repeated stages rises.

While making decisions with the help of systems techniques, decision-makers often have an operational focus on different temporal and spatial scales, and usually tend to define what is important in terms of processes they can really observe at individually characteristic scales of attention. In fact, different decision-makers may see the problem quite differently from each other, and this must be respected, especially when the decision-making process is organised in a group.

Since consequences or impacts of climate scenarios, which may occur far in the future, are to be evaluated by the decision-making tools, it is also necessary to provide tools for the detailed analysis of data, identifying sets of feasible solutions, searching for optimal or suboptimal solutions, etc. The most common classes of such instruments are shown in Figure 1. Depending on the decision-making framework and type of analysis, one or more of the listed methods and technologies may be used before or after running any regional climate model. Standard and/or

advanced methods and techniques, combined with heuristics and meta-heuristics, may help to provide an information base for decision-making either to individuals (e.g. irrigators in agriculture), or to political and other associations or larger bodies such as municipal, regional, or governmental agencies. Because climate models are inherently uncertain to a greater or lesser extent, final modelling frameworks may include intelligent algorithms and procedures such as: genetic algorithms (GAs), ant colony systems (ACSs), particle swarm optimisation (PSO), bee colony optimisation (BCO), cuckoo nest optimisation (CNO), neural networks (NNs), etc.

Decision-making related to the assessment and evaluation of regional climate scenarios can be organised in two standard ways: (1) a bottom-up approach; and (2) a top-down approach. The first is sometimes understood as an alternative-led approach, and the second as a criteria-led approach. Both coincide with the decision hierarchy and indicate that, in the first case, alternatives (options) are analysed and mutually compared, followed by criteria analysis and the final synthesis. In the second case decision-makers analyse and validate the criteria set and subsequently proceed with the analysis of alternatives and the final synthesis.

The other aspect of these two approaches is a group context where approaches must: (1) enable stakeholders' involvement and identify options (e.g. climate scenarios, adaptation options, organisational and special measures,

SYSTEMS ANALYSIS METHODS

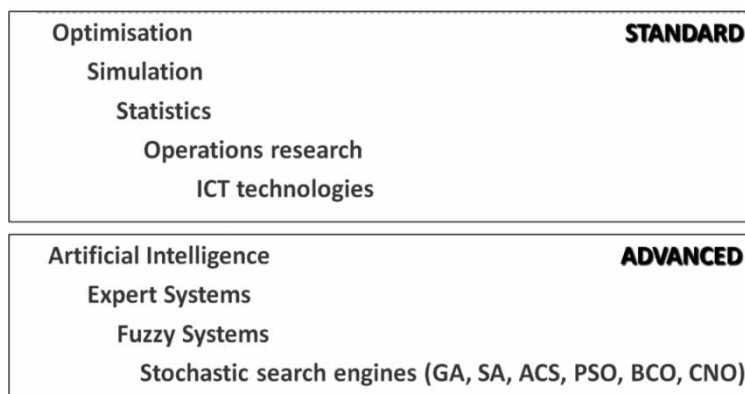


Figure 1 | Main classes of systems analysis methods used to support decision-making processes. GA – genetic algorithms; SA – simulated annealing; ACS – ant colony systems; PSO – particle swarm optimisation; BCO – bee colony optimisation; CNO – cuckoo nest optimisation.

etc.); (2) perform an evaluation of options and determine their weights as cardinal information, or priorities as ordinal information; and (3) declare the result by ordering options according to their weights or other types of scores indicating their ranks. Regarding the last action (3), weights (and corresponding ranks) can be obtained using the analytic hierarchy process (AHP), while direct ranking can be derived using methods such as outranking methods ELECTRE and PROMETHEE, or ideal-point methods Compromise Programming (CP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (for more details, see 'Preliminaries' below).

PRELIMINARIES

The analytic hierarchy process

Pairwise comparisons in AHP

In the AHP method, a decision-maker (DM) evaluates the elements at several levels of the problem hierarchy by making pairwise comparisons of the decision elements (criteria, sub-criteria, alternatives). The DM iteratively carries out a face-to-face comparison of all pairs of decision elements at a given level of the hierarchy, with respect to each element in the upper level. To accomplish this, a 9-point ratio scale (Table 1) is usually used. Comparisons made between all elements at a given level versus one element in the upper level are inserted into a corresponding comparison matrix as upper triangle entries; lower triangle entries are symmetrical reciprocals of entries in the upper triangle. Entries at the main diagonal are equal to one,

since each element compared with itself is equally important. Once the comparison matrix is completed, the selected matrix calculus or optimisation method is used to derive local priorities represented by weights of paired decision elements, with respect to superior decision elements in the upper level (for instance, weights of sub-criteria with respect to a given criterion). Once all local vectors are computed, AHP performs a synthesis and derives the final weights of elements in the bottom level vs. the goal at the top level.

A standard AHP hierarchy has three levels positioned (from the top down) as a goal, criteria set, and alternatives set. If a complete hierarchy is assessed in the described way, it is considered to be a standard AHP. If only two levels exist, for instance a goal and a set of criteria, then only one comparison matrix is created by the DM, and only one application of the prioritisation method is necessary to derive the priority vector w (containing the weights of the criteria with respect to the stated goal). This case embodies the 'spirit of AHP', rather than AHP itself, because there is no synthesis which is typical for standard AHP; note that the same is true for any local prioritisation within standard AHP.

Most papers related to the AHP framework deal with a single (local) matrix, and develop theory and discussion without going into the issue of synthesis and the implications of various properties of single (local) matrices, such as their representativeness, consistency, and other issues. In group AHP applications, the importance of consistency emerges and various approaches may be used to derive correct group priority vectors.

Before a brief summary of important preliminaries is presented, it should be recalled that, for many reasons, and especially if the number of paired elements is larger than four, the DM rarely makes perfectly consistent judgments, due to many reasons (for example, because of a lack of knowledge, or uncertainty in the environment). Matrix $A(a_{ij})$ in which the judgments are inserted is rarely consistent. That is, it rarely satisfies the transitive condition $a_{ij} = a_{ik} a_{kj}$ for all $i, j, k = 1, 2, \dots, n$. Number n denotes number of decision elements used in the pairwise comparison, i.e. the size of a matrix. In fact, the exact values $a_{ij} = w_i/w_j$ will occur only if the matrix is perfectly consistent, which means that $a_{ij} = w_i/w_j$ for all i and j .

Table 1 | Relative importance scale (Saaty 1980)

Definition	Assigned value
Equally important	1
Weak importance	3
Strong importance	5
Demonstrated importance	7
Absolute importance	9
Intermediate values	2,4,6,8

Prioritisation methods

Because the DM's evaluations a_{ij} are very rarely perfect, and the transitive rule is frequently violated, the comparison matrix A is inconsistent and its elements are only approximations of the derived priority vector w , i.e. $a_{ij} \approx w_i/w_j$. The main reason why the prioritisation method must be used to estimate vector w is because it is not unique, due to inconsistencies. To measure the quality of the estimation process, certain measures are commonly used either as measures associated with specific methods, or as general error measures applicable to all known methods. Note that here the term 'error' actually means 'inconsistency'.

It is worth mentioning that deriving the priorities of the compared elements is a critical issue, because prioritisation techniques produce different results with different deviations from elicited judgments. Prioritisation methods can generally be classified into two categories: (1) matrix calculus methods; and (2) optimisation methods as given in Srdjevic (2005). The performance of prioritisation methods can be measured using different evaluation criteria. Some measures are method-specific, but two are more general and applicable to all methods: total Euclidean distance (ED) and minimum violations (MV). ED and MV are error measures, both use different metrics, but they share the common property of measuring the accuracy of the prioritisation method and determining the ranking properties. Many researchers use these evaluation measures to check the quality of prioritisation methods in both individual and group AHP decision-making applications.

Additive normalisation (AN) solution

The desired priority vector w is obtained in two simple steps. Firstly, the elements of each column of matrix A should be divided by the sum of that column. Secondly, elements in each resulting row are added and finally row sums are divided by the number of elements in the row. The procedure is described by Equations (1) and (2):

$$a'_{ij} = \frac{a_{ij}}{\sum_{i=1}^n a_{ij}}, \quad i, j = 1, 2, \dots, n \quad (1)$$

$$w_i = \left(\frac{1}{n}\right) \sum_{j=1}^n a'_{ij}, \quad i = 1, 2, \dots, n. \quad (2)$$

The AN method in some examples outperformed more sophisticated methods, especially if A is close to consistent as given in Srdjevic (2005). Kou & Lin (2014; p. 229) added that AN 'can be viewed as an approximation to eigenvector method, proposed by Saaty (1980) in practice'.

Consistency and distance measures

Three consistency measures are commonly used in standard AHP applications if additive normalisation is applied as the prioritisation method to extract the priority vector w from pairwise comparison matrix A .

- *Consistency ratio.* The first measure is the consistency ratio (CR) (Saaty 1980), which is a part of the standard AHP procedure. It was suggested by Saaty (1980) that the maximum level of the DM's inconsistency should be 0.10. However, in different studies (e.g. Karlsson 1995) CR values up to 0.15 are also accepted as tolerable inconsistency.
- *Euclidean distance.* The universal L_2 error metrics and measure of consistency in AHP is the total ED between all judgment elements a_{ij} in the comparison matrix, and related ratios of the priorities contained in the derived vector w :

$$ED = \left[\sum_{i=1}^n \sum_{j=1}^n (a_{ij} - w_i/w_j)^2 \right]^{1/2}. \quad (3)$$

- *Minimum violation.* The minimum violation criterion (MV) sums up all violations associated with the priority vector w :

$$MV = \sum_{i=1}^n \sum_{j=1}^n I_{ij} \quad (4)$$

where

$$I_{ij} = \begin{cases} 1 & \text{if } w_i > w_j \text{ and } a_{ji} > 1 \\ 1 & \text{if } w_i < w_j \text{ and } a_{ji} < 1 \\ 0.5 & \text{if } w_i = w_j \text{ and } a_{ji} \neq 1 \\ 0.5 & \text{if } w_i \neq w_j \text{ and } a_{ji} = 1 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

The ‘conditions of violation’, defined by (5), penalise possible order reversals such as this: if alternative j is preferred to alternative i (i.e. $a_{ji} > 1$), but the derived priorities are such that $w_i > w_j$, then there is a ‘violation’, or element preference reversal.

Notice that in cases when more matrices of different sizes are analysed, normalisation of consistency measures is advisable to preserve consistent comparisons, for instance, by dividing values for ED and MV by n^2 ; in the application described in the next section this was not necessary.

Spearman’s rank correlation coefficient. Decision elements of comparison matrix A can be ranked if the priority vector $w = (w_1, w_2, \dots, w_n)$ is computed. In a group context, if there is a group priority vector with related ranks of the elements, then it can be considered to be the reference vector for each participating DM, and Spearman’s rank correlation coefficient can be calculated as:

$$S = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad (6)$$

D_i is a rank difference between the rank of the element from vector w , and the rank of the corresponding element in the reference vector, while n represents the number of ranked elements. The coefficient S describes the positive or negative correlation between vector w and the reference vector, and can have a value in the range $[-1, 1]$; S has a value of -1 if the elements in w and the reference vector have opposite ranks (ideal negative correlation); the value $+1$ of coefficient S shows that the elements are fully matched (ideal positive correlation); if $S = 0$, the ranks do not correlate.

In group decision-making applications, the Spearman’s rank correlation coefficient can be understood as a statistical distance measure, or statistical conformity of ranks obtained by the decision-makers with reference ranks for an ‘average’ DM. The use of Spearman’s rank correlation coefficient to measure conformity of an individual’s opinions to those of the group average assumes that the number of decision-makers is not too small. Namely, ranking is a relative measure, not an absolute measure, and this means that two items can differ

significantly in rank preference but be relatively close in absolute preference. In the case of larger groups, possible misguidance is eliminated.

Borda count (BC) and Approval voting (AV)

Typical social choice theory methods are known as the preferential and non-preferential voting methods. The two methods used in this research are the Borda count and Approval voting. They will be briefly presented here using the same terminology as in D’Angelo *et al.* (1998) and Srdjevic (2007).

Borda count

In applying this preferential voting method, a preference $m \times n$ table is usually constructed, where m is the number of voters and n is the number of options to be voted upon. Each row represents the ranking of the options performed by one voter. If j is the best option for voter i , then the rank number is $r_{ij} = 1$; if j is the second-best option, then $r_{ij} = 2$, and so on; if option j is the worst, then rank number is $r_{ij} = n$. This way, each row of the preference table is a permutation of the integers $1, 2, \dots, n$. The option with the lowest point total wins the election, and is declared to be the social choice.

Approval voting

In contrast to the Borda count, which uses information directly from a preference schedule, Approval voting does not do so. It is non-preferential because voters can vote for as many options as they choose. Each approved option receives one vote and the candidate with the most votes wins.

CASE STUDY

Problem formulation and methodology

A need for a robust decision process implies the selection of a scenario which meets intended goals – e.g. to select an appropriate model for simulating global, regional, or local climate processes; to recognise impacts of priorities

on allocation and access to safe water; to reduce floods and all kinds of damages; to increase reliability and resilience of water and other infrastructural systems; and to do it across a variety of plausible futures represented by various scenarios. A common assumption in methodologies devoted to climate scenario evaluation is that policy-relevant impact assessments related to future climate change must recognise different types of climate scenarios, such as those described by outputs from climate models, incremental scenarios useful in sensitivity studies, analogue scenarios, and a variety of other general scenarios.

Assessment of climate change scenarios can be performed in different ways and with different mathematical, methodological, and organisational tools. One promising possibility is explored and described here in more detail. A problem is formulated in order to evaluate relevant criteria which can be used in global, regional, and local assessments of climate scenarios. An approach is developed as a group decision-making methodology to weight by importance a set of five criteria, applicable in any policy-relevant impact assessment.

As suggested in Smith & Hulme (1998), and without reduction in generality, the following five criteria are used for evaluating the suitability of climate scenarios:

C1 – Consistency at a regional level with global projections.

Scenario changes in regional climate may lie outside the range of global mean changes, but should be consistent with theory and model-based results (CONS).

C2 – Physical plausibility and realism. Changes in climate should be physically plausible, such that changes in different climatic variables are mutually consistent and credible (PLAU).

C3 – Appropriateness. Appropriateness of information for impact assessments. Scenarios should present climate changes at an appropriate temporal and spatial scale, for enough variables, and over an adequate time horizon, to allow for impact assessments (APPR).

C4 – Representativeness. Representativeness of the potential range of future regional climate change (REPR).

C5 – Accessibility. The information required for developing climate scenarios should be readily available and easily accessible for use in impact assessments (ACCE).

To validate characteristics of climate scenarios in a multi-criteria sense it is necessary to define the importance of each criterion, and to specify ratings of scenarios with respect to the criteria set. The following decision-making problem is stated accordingly as:

Rank by importance criteria C1-C5 for evaluating the suitability of climate scenarios for the use in policy-relevant impact assessment.

The AHP methodology of performing pairwise comparisons and prioritisation was adopted to elicit judgments from 16 PhD students, participants of the training school 'Validating Regional Climate Projections'. (The school took place in Trieste in Italy in late 2015, within the framework of the COST Action ES1102 project.) Students were identified as decision-makers DM1 to DM16. Participants acted as a unique group, but without sharing information. Each participant individually performed pairwise comparisons of criteria by using Saaty's 9-point ratio scale (Table 1). Once judgments were completed, prioritisation of criteria was performed for each participant, and the consistency of judgments was individually controlled. The aggregation of computed priority vectors produced unique group output, i.e. a group priority vector as a decision represented by weights, and related ordering of criteria by importance for further evaluations of climate scenarios.

The decision-making process was organised in the following way. On the first day, students were briefed on decision-making processes, AHP, the Borda count, and Approval voting methods. The next day, in two separate sessions, students did an exercise as described below.

During the *morning session* (the first part of the decision-making process), the following activities were performed:

1. A detailed evaluation sheet was distributed to participants and instructions were given on how to assess options and perform judgments by pairwise comparisons as required by the AHP method, and then to 'vote criteria' by Borda count and Approval voting.
2. Participants performed pairwise comparisons and voted within 20 minutes (individual paperwork with evaluation sheets and ballots).

3. Evaluation sheets and ballots were collected and coded for computation.
4. AHP priority vectors (weights of criteria) were computed for all individuals and geometrically aggregated.
5. Demonstrated consistencies (CRs, EDs, and MVs) of participants were analysed.
6. Based on individual weights and ranks of criteria, Spearman's rank correlation coefficient was computed based on the reference ranking of criteria in the aggregated vector.
7. Ballots were collected and the votes were analysed using the Borda count and Approval voting methods.

The *afternoon session* consisted of the following activities:

1. Participants were (randomly) divided into three groups: G1, G2, and G3 with five, six, and five members, respectively.
2. One evaluation sheet, the same used in the morning session, was distributed to groups, and instructions were repeated to participants on how to perform pairwise comparisons. Discussion within each group was advised (and during the session strongly enhanced) in order to produce efficient brainstorming and achieve a group consensus on judgments, while making pairwise comparisons of criteria.
3. Each group produced a unique group pairwise comparison matrix within approximately 30 minutes.
4. Three evaluation sheets were collected and judgments coded for computation.
5. Priority vectors obtained by each group were geometrically aggregated. Demonstrated group consistencies (CRs, EDs, and MVs) were analysed and compared with values obtained individually during the morning session.
6. Ballots were delivered to participants to perform, firstly, preferential voting using the Borda count and, secondly, non-preferential voting using the Approval voting method. Ballots were collected and the results summarised and interpreted.

Grand summaries of the results obtained during the two sessions were discussed with the participants and final conclusions were derived.

Results and discussion

The morning session

All judgments made by PhD students during the first part of the exercise are presented in [Table 2](#). The results of prioritisation performed with the AN method (Equations (1) and (2)) and consistency measures CR, ED, and MV are summarised in [Table 3](#).

The participants validated criteria differently and demonstrated a wide range of consistency, from very inconsistent (e.g. for DM3 and DM1 the computed CR measures were 0.853 and 0.470, respectively), to very consistent (e.g. for DM5 and DM4 the values of CR were 0.032 and 0.042, respectively). On average, the group was on the edge of something that could be accepted as a frontier consistency of, say, 0.200, which is double the suggested tolerance value of 0.100 ([Saaty 1980](#)). At this point, there is not known to be any evidence in the literature about either tolerable inconsistency for the group, or how consistency could be correlated with the size of the group. Therefore, an average value of 0.194 for CR is accepted as reasonable and, furthermore, the group weights of criteria shown in [Table 3](#) are declared to be representative for the group.

The total Euclidean distances computed using Equation (3) are generally low, with an average value of 6.382 with 10 DMs (or 62%) below that value. The correlation coefficient for the measures CR and ED is 0.78, which is statistically justified based on a sample of size 16 (equal to the number of DMs).

The consistency measure MV, which indicates the degree of violations associated with values of elicited judgements a_{ij} and entries w_i and w_j in the computed priority vector for a given DM, varies from zero to three. Our earlier experiments show that, for a comparison matrix of size five, the value of $MV = 2$ can be accepted if CR and ED are not out of tolerance boundaries ([Srdjevic & Srdjevic 2013](#)). In this case about half of all DMs are slightly above or below the value $CR = 0.10$, and because the correlation of 0.78 between CR and ED is satisfactory, it can be concluded that the average value $MV = 1.38$ is also satisfactory; notice that for six DMs no violations are identified (i.e. $MV = 0$).

Aggregated weights show that the top ranked criterion is C2, with the final weight $w_2 = 0.305$. This means that the

Table 2 | Judgments made by individual decision-makers

Judgments from the scale ^a										Decision-makers
7 ^b	5	3	1	3	5	3	4	1/3	3	DM1
1/5	2	1/3	1/5	3	5	2	3	1	1/5	DM2
5	9	7	5	3	1/7	1/5	7	9	1	DM3
3	1	3	4	1	4	3	5	5	1	DM4
1/3	2	3	5	5	4	6	2	3	2	DM5
5	3	1	7	5	1/7	3	1/7	1	7	DM6
1/7	5	2	5	5	3	6	1/4	3	4	DM7
1/3	1/5	1/3	3	1	1	3	3	3	1	DM8
5	3	1/3	5	1/5	1	3	3	5	3	DM9
1/3	5	5	7	3	5	7	3	5	3	DM10
1/5	1/7	1/4	3	1/6	6	5	8	9	2	DM11
1/5	1/2	1/2	1/3	5	2	1	2	3	1	DM12
1/2	5	5	8	7	3	8	2	5	3	DM13
1/5	7	7	5	8	9	7	3	1/4	1/6	DM14
3	1/5	3	5	1/3	1/3	1	2	3	5	DM15
1/5	1/3	1/3	5	1	3	7	3	5	5	DM16

^aNumerical values from the nine-point scale in Table 1 are taken from the upper triangle in each matrix, row by row in a top-down direction.

^bE.g. value 7 corresponds to judgment a_{12} in the matrix created by the decision-maker DM1 (meaning that criterion C1 is much more important than criterion C2).

Table 3 | AHP weights and ranks of options obtained for individual participants

Decision-makers	Weights of criteria					Consistency measures		
	C1	C2	C3	C4	C5	CR	ED	MV
DM 1	0.383	0.229	0.107	0.121	0.160	0.470	8.051	3.0
DM 2	0.100	0.401	0.151	0.091	0.256	0.203	4.775	3.0
DM 3	0.460	0.085	0.230	0.121	0.105	0.853	13.822	2.0
DM 4	0.335	0.218	0.306	0.072	0.069	0.042	2.795 ^a	3.0
DM 5	0.235	0.482	0.136	0.093	0.054	0.032 ^a	3.808	0.0 ^a
DM 6	0.340	0.132	0.062	0.417	0.050	0.108	6.128	2.0
DM 7	0.217	0.487	0.079	0.172	0.045	0.139	7.596	0.0 ^a
DM 8	0.113	0.259	0.355	0.176	0.097	0.113	3.579	3.0
DM 9	0.321	0.104	0.302	0.228	0.046	0.271	5.850	2.0
DM 10	0.313	0.425	0.148	0.076	0.038	0.097	5.894	0.0 ^a
DM 11	0.068	0.232	0.560	0.099	0.040	0.172	8.422	0.0 ^a
DM 12	0.069	0.375	0.228	0.141	0.187	0.144	4.352	2.0
DM 13	0.330	0.428	0.115	0.092	0.036	0.083	6.462	0.0 ^a
DM 14	0.253	0.533	0.057	0.032	0.125	0.151	10.123	0.0 ^a
DM 15	0.259	0.080	0.0386	0.204	0.072	0.156	5.637	1.0
DM 16	0.107	0.367	0.317	0.167	0.042	0.074	4.825	1.0
Aggregated weights	0.248	0.305	0.214	0.146	0.086	Averages		
Ranks	(2)	(1)	(3)	(4)	(5)	0.194	6.382	1.38

Best values of consistency measures (by individual).

group believes that in evaluating climate scenarios, it is most important to ensure that the changes in climate are physically plausible. Furthermore, the group thinks that any model used for evaluating climate scenarios should consist of mathematical and software mechanisms which conceptually ensure that at least major changes in climatic variables are mutually consistent and credible.

The second top-ranked criterion is C1, with group weight $w_1 = 0.248$. The group opinion is that consistency at a regional level with global projections is of high importance in evaluating the suitability of climate scenarios for use in policy-relevant impact assessments. Although scenario changes in regional climate may lie outside the range of global mean changes, any model used for the evaluation of suitability of climate scenarios should ensure that its output is consistent with theory.

It is interesting to note that the accessibility criterion (C5) is positioned last, with a weight less than 10% ($w_5 = 0.086$). The group thinks that for evaluating the quality of models that will perform impact assessments, the least important criterion is the availability and accessibility of the information required for developing (modelling) climate scenarios.

The AHP results are put into the social choice theory context by applying the Borda count as shown in Table 4; the criteria received Borda points after they were ranked by weight (one for the largest weight, and five for the smallest; Cf. Table 3).

Once all points in Table 4 are summarised, it is easy to see that criteria ranking is the same if cardinal (AHP) and ordinal (Borda) information is used (Cf. ranks in the last rows of Tables 3 and 4).

The ordinal information in Table 4 is directly used to compute the Spearman correlation coefficient for each DM, i.e. to see how much each individual ordering differed from the group ordering given in the last row of Table 4. Note that three decision-makers (DM5, DM10, and DM13) ranked criteria in the same way as the complete group did, i.e. their Spearman correlation coefficient was at the maximum value of 1. In only one case (DM15) the value S was negative, and this DM considered criterion C1 the least important, completely opposite to the rest of the group. The average value of the Spearman correlation coefficient for the group is $S = 0.54$, and this positive value shows

Table 4 | Criteria ranks (from AHP) in the Borda count context and their statistical deviation from the ranks derived from geometrically aggregated weights

Decision-maker	Ranks of criteria					Spearman corr. coef. S
	C1	C2	C3	C4	C5	
DM 1	1	2	5	4	3	0.5
DM 2	4	1	3	5	2	0.3
DM 3	1	5	2	3	4	0.0
DM 4	1	3	2	4	5	0.7
DM 5	2	1	3	4	5	1.0 ^a
DM 6	2	3	4	1	5	0.3
DM 7	2	1	4	3	5	0.9
DM 8	4	2	1	3	5	0.5
DM 9	1	4	2	3	5	0.4
DM 10	2	1	3	4	5	1.0 ^a
DM 11	4	2	1	3	5	0.5
DM 12	5	1	2	4	3	0.3
DM 13	2	1	3	4	5	1.0 ^a
DM 14	2	1	4	5	3	0.7
DM 15	2	5	1	3	4	-0.1
DM 16	4	1	2	3	5	0.6
Sum	39	34	42	56	69	
Ranks of criteria in AHP aggregated vector	(2)	(1)	(3)	(4)	(5)	0.54

Best values (i.e. individual ranking is equal to the group ranking).

that members of the group generally (not perfectly, but satisfactorily) agreed with the common group ranking.

The afternoon session

The afternoon session consisted of two parts. In the first part participants were divided into three groups. After discussion within each group, they produced the pairwise comparison matrices presented in Table 5.

Prioritisation was applied to the matrices from Table 5, and the aggregation of priority vectors for the groups produced the results shown in Table 6. Criterion C2 again received the highest weight ($w_2 = 0.417$), which is higher than the weight obtained in the morning session when individual priority vectors for all DMs were aggregated (see $w_2 = 0.305$ in Table 3). An interesting result is that criteria C1 and C3 changed places – C1 was weighted with half the value from the morning session (0.128 vs. 0.248) and C3 received a 35% higher weight (0.292 vs. 0.214) and was

Table 5 | Judgments made by consensus in the three groups

Judgment from the scale ^a										Group
1/3	5 ^b	7	7	5	7	9	3	4	4	G1: DM5, DM6, DM7, DM8, DM13
1/4	1/5	1	3	1	3	7	4	9	3	G2: DM1, DM9, DM10, DM11, DM12, DM15
1/5	1/5	1/4	1/3	1	4	3	7	5	1/3	G3: DM2, DM3, DM4, DM14, DM16

^aNumerical values from the nine-point scale in Table 1 are taken from the upper triangle in each matrix, row by row in the top-down direction.

^bE.g. value 5 corresponds to judgment a_{13} in the matrix created by the group G1 (meaning that criterion C1 is much more important than criterion C3).

Table 6 | Criteria weights obtained in a sub-group context

Group	C1	C2	C3	C4	C5	CR	ED	MV
G1: DM5, DM6, DM7, DM8, DM13	0.301 (2)	0.481 (1)	0.113 (3)	0.071 (4)	0.034 (5)	0.101	7.253	0.00 ^a
G2: DM1, DM9, DM10, DM11, DM12, DM15	0.099 (4)	0.347 (2)	0.404 (1)	0.109 (3)	0.041 (5)	0.009 ^a	2.178 ^a	0.08
G3: DM2, DM3, DM4, DM14, DM16	0.051 (5)	0.321 (2)	0.401 (1)	0.094 (4)	0.133 (3)	0.095	5.403	0.04
Aggregated weight and ranks	0.128 (3)	0.417 (1)	0.292 (2)	0.100 (4)	0.063 (5)			
Average value of consistency measures						0.068	4.945	0.04

^aBest values.

ranked accordingly, i.e. as the second most important. Grouping individuals 'enhanced' the importance of criterion C2 (physical plausibility and realism), pushed on the second-place criterion C3 (appropriateness of information for impact assessments) and placed criterion C1 in the third position (consistency at regional level with global projections).

The results in Table 6 show that, after a break between the morning and afternoon sessions, participants in the decision-making exercise 'evolved' in how they evaluated the importance of criteria during the afternoon session. They had more time to discuss and reach a consensus while comparing criteria, which resulted in more consistent judgments within groups, and much better values of consistency measures CR, ED, and MV (Cf. Table 6). Note that the groups demonstrated different preferences for criteria, and that G2 was almost perfectly consistent (CR = 0.009) with their ranking of criteria (C3-C2-C4-C1-C5) which differs from the aggregated ranking (C2-C3-C1-C4-C5).

The second part of the afternoon session was conducted after a short break. Two ballots with descriptions of five criteria were delivered to participants so that they could individually rank criteria. The first ballot was for preferential voting in a Borda count manner, i.e. by giving one point to top ranked criterion, two points to the second, and so on, until the last criterion, which receives five points.

The 16 voters produced the results shown in Table 7; note that in cases when the voter did not want to give

Table 7 | Criteria ranking obtained by Borda count voting method (preferential)

Voter (Decision-maker)	Criteria				
	C1	C2	C3	C4	C5
DM 1	2	1	3	4	5
DM 2	4	1	2.5	2.5	5
DM 3	2	3	1	4	5
DM 4	2	3	1	4	5
DM 5	2	1	3	4	5
DM 6	1	5	3	2	4
DM 7	2	1	4	3	5
DM 8	4	2	1	3	5
DM 9	3	4	1	2	5
DM 10	2	1	3	4	5
DM 11	3	2	1	4	5
DM 12	3.5	3.5	5	1.5	1.5
DM 13	2	1	3	4	5
DM 14	2	1	4	5	3
DM 15	2	3	1	4	5
DM 16	3	1	2	4	5
Number of points	39.5	33.5	38.5	55	73.5
Rank	(3)	(1)	(2)	(4)	(5)

preferences or shared ranks, the number of points was divided accordingly.

The second ballot was used by participants to simply approve criteria they considered important, but this time without giving preferences. Participants were instructed to tick any number of criteria, assuming that the minimum is one and maximum is all five criteria. The results are shown in Table 8.

Table 9 summarises all results. According to all applied procedures, group opinion is that the most important criterion in assessing the quality of models for evaluating climate scenarios is criterion C2 (physical plausibility and realism), followed by criteria C3 (appropriateness) and C1 (consistency at regional level with global projections).

Participants also thought that, in the assessment, climate models' representativeness of the potential range of future REPR (C4) and accessibility of the information required for developing climate scenarios (C5), play a less important role than previously described characteristics of models measured via criteria C2, C3, and C1.

Table 8 | Criteria ranking obtained by Approval voting method (non-preferential)

Voter (Decision-maker)	Criteria				
	C1	C2	C3	C4	C5
DM 1		+			
DM 2	+	+	+		+
DM 3	+		+	+	
DM 4	+	+	+	+	+
DM 5	+	+	+	+	+
DM 6	+	+		+	
DM 7	+			+	
DM 8		+	+		
DM 9			+	+	
DM 10		+			
DM 11			+		+
DM 12		+	+		+
DM 13	+	+	+	+	+
DM 14	+	+			+
DM 15	+		+	+	
DM 16	+	+	+	+	
Number of approvals	10	11	11	9	7
Rank	(3)	(1-2)	(1-2)	(4)	(5)

Table 9 | Grand summary: Borda points for all applied procedures during morning and afternoon sessions

Decision-making procedures/Ranks	Criteria				
	C1	C2	C3	C4	C5
AHP (altogether) (Table 2)	(2)	(1)	(3)	(4)	(5)
AHP (3 groups) (Table 5)	(3)	(1)	(2)	(4)	(5)
Borda count (altogether) (Table 6)	(3)	(1)	(2)	(4)	(5)
Approval voting (altogether) (Table 7)	(3)	(1.5)	(1.5)	(4)	(5)
Sum of Borda points	(11)	(4.5)	(8.5)	(16)	(20)
Final ranking	(3)	(1)	(2)	(4)	(5)
THE FINAL RANKING OF CRITERIA					
C1 Consistency at regional level with global projections				THIRD	
C2 Physical plausibility and realism				BEST	
C3 Appropriateness				SECOND BEST	
C4 Representativeness				FOURTH	
C5 Accessibility				FIFTH	

CONCLUSIONS

There is scientific agreement that there are undesirable changes in the global climate which are easily recognised in the general trends of changing precipitation patterns, greenhouse gas emission impacts, melting of ice caps, and frequent meteorological droughts and other intermittent extreme weather events such as flooding. The prognosis and evaluation of the impacts of possible climate scenarios for the future are becoming a permanent concern for society worldwide.

Most of the existing models representing climate change impacts inherently represent real uncertainty since they are all based on incomplete and imperfect knowledge. Still, there is no general agreement about which critical (minimal) requirements models must fulfil with respect to data input, confidence level of parameters, and up- (or down-) scaling metrics applied. The main issue in using models is how much the analyst or end-user (for instance, a stakeholder) of the model really trusts the model. In most cases the value (quality) of the model is 'measured' by its ability to reproduce the present climate and project future climate change; for instance, probabilities of future climate changes can be computed with the statistical models, but can also

be derived as projections from climate models. To ensure good projections, models must be calibrated and verified by using climate data from different historical periods, sometimes with data gaps, 'white noise', etc. It is correctly stated in many scientific reports that validation of any model can be viewed simultaneously as a subjective and objective (expert judgment) process, depending on which will prevail. This, furthermore, means that validation is subject to manipulation; positive, negative, or neutral. In any case, validation of several different models for climate changes prognosis and determining the impacts of different climate scenarios remains an active research topic in the scientific community, especially because of the perceived utility of these models for decision-making.

Most decisions in environmental management come with long-term commitments, necessary to account for the risks of failures of infrastructure, flood events, meteorological droughts, etc. Decisions such as the adoption of flood mitigation plans may have consequences over periods of 50–200 years, and such decisions can be vulnerable to changes in climate and land use conditions (Yang et al. 2012). Therefore, to provide a reliable framework for decision-making about which model to adopt, it is necessary to create a set of relevant criteria which will enable the assessment of models and, finally, the identification of the most appropriate one. The decision-making process can be organised and manipulated in many ways, and supported by one or more mathematical/software tools, which are sometimes quite conceptually different. It is extremely important to preserve the application of multi-criteria methods which offer clarity and feedback to the user such that the model structure is not a black-box and the user can conveniently examine how changes in value judgments affect the decision made at the end of a process (Bell et al. 2003).

This paper demonstrates how two distinct methodologies can be combined for determining the importance of criteria commonly used in validating the quality of regional climate models aimed at exploring the impacts of climate change scenarios. The AHP methodology is used to compute the weights of five selected criteria by a group of 16 PhD students. The students also used two other methods from the social choice theory framework: the Borda count and Approval voting. They voted for criteria in a preferential and non-preferential manner, respectively.

The main research idea was to check how real decision-makers (voters) behave if they act as individuals and/or members of a group, and what results distinct methodologies may produce when evaluating the criteria set required in further assessments of possible future scenarios of climate changes at regional scales. The results are such that one can trust in results obtained by either multi-criteria decision-making or social choice theory methods or their combination. If carefully led, the process of decision-making or voting (or both during the same sessions) produce similar results, at least for the top-ranked options (criteria), and can therefore be declared to be trustworthy.

SOFTWARE

All computations described in this paper are performed with our original software DECIDE consisting of 3,620 code lines in seven separate programs with special routines for data interchange. Fortran programming language is used and DECIDE software is not available for distribution.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education, Science and Technological Development of Serbia under the grant 174003 (2011–2016) – Theory and application of analytic hierarchy process (AHP) in multi-criteria decision making under conditions of risk and uncertainty (individual and group context). In part it is also supported by the COST Action ES1102 – Validating and Integrating Downscaling Methods for Climate Change Research (VALUE).

REFERENCES

- Bell, M. L., Hobbs, B. F. & Ellis, H. 2003 [The use of multi-criteria decision-making methods in the integrated assessment of climate change: implications for IA practitioners](#). *Socio-Economic Planning Sciences* **37**, 289–316.
- D'Angelo, A., Eskandari, A. & Szidarovszky, F. 1998 [Social choice procedures in water-resource management](#). *Journal of Environmental Management* **52** (3), 203–210.
- Feser, F., Rockel, B., von Storch, H., Winterfeldt, J. & Zahn, M. 2011 [Regional Climate Models add Value to Global Model](#)

- Data: a Review and Selected Examples. In: *American Meteorological Society*, September 2011, pp. 1181–1192.
- Giorgi, F. & Mearns, L. O. 1999 [Introduction to special section: regional climate modelling revisited](#). *Journal of Geophysical Research* **104**, 6335–6352.
- Giorgi, F., Jones, C. & Asrar, G. R. 2009 Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bulletin* **58** (3), July 2009.
- Hallegatte, S., Ankur, S., Brown, C., Lempert, R. & Gill, S. 2012 *Investment Decision Making Under Deep Uncertainty – Application to Climate Change*. World Bank Policy Research Working Paper No. 6193. Available at: <https://ssrn.com/abstract=2143067> (accessed 24 April 2018).
- Karlsson, J. 1995 *Towards a Strategy for Software Requirements Selection Licentiate*. PhD thesis 513, Linköping University, Sweden.
- Kou, G. & Lin, C. 2014 [A cosine maximization method for the priority vector derivation in AHP](#). *European Journal of Operational Research* **235**, 225–232.
- Laprise, R., De Elia, R., Caya, D., Biner, S., Lucas-Picher, P. H., Diaconescu, E., Leduc, M., Alexandru, A. & Separovic, L. 2008 [Challenging some tenets of regional climate modelling](#). *Meteorology and Atmospheric Physics* **100**, 3–22.
- Liu, Y., Gupta, H., Springer, E. & Wagener, T. 2008 [Linking science with environmental decision making: experiences from an integrated modelling approach to supporting sustainable water resources management](#). *Environmental Modelling & Software* **23**, 846–858.
- Maraun, D., Widmann, M. J., Gutiérrez, M., Kotlarski, S., Chandler, R. E., Hertig, E., Wibig, J., Huth, R. & Wilcke, R. A. 2015 [VALUE: A framework to validate downscaling approaches for climate change studies](#). *Earth's Future* **3**, 1–14.
- Saaty, T. L. 1980 *The Analytic Hierarchy Process*. McGraw-Hill, New York, NY.
- Seneviratne, S. I. 2012 *Changes in Climate Extremes and Their Impacts on the Natural Physical Environment, ETH Zurich, Switzerland*. http://ipcc.ch/pdf/presentations/SREX_geneva_march2012/ipcc_srex_seneviratne_geneva_201203.pdf (accessed 24 January 2017).
- Smith, J. B. & Hulme, M. 1998 Climate change scenarios. In: *Handbook on Methods of Climate Change Impacts Assessment and Adaptation Strategies: Version 2.0* (J. Feenstra, I. Burton, J. B. Smith & R. S. J. Tol, eds). UNEP/IES, Amsterdam.
- Srdjevic, B. 2005 [Combining different prioritization methods in analytic hierarchy process synthesis](#). *Computer & Operations Research* **32**, 1897–1919.
- Srdjevic, B. 2007 [Linking analytic hierarchy process and social choice methods to support group decision-making in water management](#). *Decision Support Systems* **42** (4), 2261–2273.
- Srdjevic, B. & Srdjevic, Z. 2013 [Synthesis of individual best local priority vectors in AHP-group decision making](#). *Applied Soft Computing* **13**, 2045–2056.
- Thiemeßl, M. J., Gobiet, A. & Heinrich, G. 2012 [Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal](#). *Climatic Change* **112**, 449–468.
- Wang, Y., Leung, L. R., McGregor, J. L., Lee, D., Wang, W., Ding, Y. & Kimura, F. 2004 [Regional climate modelling: progress challenges and prospects](#). *Journal of the Meteorological Society of Japan* **82**, 1599–1628.
- Yang, J.-S., Chung, E.-S., Kim, S.-U. & Kim, T.-W. 2012 [Prioritization of water management under climate change and urbanization using multi-criteria decision-making methods](#). *Hydrology and Earth System Sciences* **16**, 801–814.

First received 27 September 2017; accepted in revised form 21 April 2018. Available online 7 May 2018